УДК 004.8

Поддержка принятия решений на основе адаптивных нечетких ситуационных сетей и глубокого обучения с подкреплением

Шапкин А. П., Борисов В. В.

***Постановка задачи**: известные нечеткие ситуационные модели позволяют представлять процесс поддержки решений в виде сети ситуаций и управляющих воздействий, однако недостаточно эффективны в условиях неопределенности из-за необходимости учета изменения системных и внешних факторов, стратегий принятия решений и иных ограничений. **Цель работы**: реализация поддержки принятия решений на основе предлагаемой разновидности адаптивных нечетких ситуационных сетей с использованием глубокого обучения с подкреплением, обеспечивающего оценку допустимых управляющих воздействий с учетом их долгосрочных последствий. **Используемые методы**: для построения сетевой структуры модели, оценки и выбора последовательности управляющих решений используется глубокое обучение с подкреплением. Для стабилизации процесса обучения с подкреплением применяются глубокая рекуррентная нейронная сеть. Недопустимые действия исключаются посредством бинарной маски переходов, формируемой по текущей структуре адаптивной нечеткой ситуационной сети. Стратегия управления задается параметризованной функцией награды, включающей компоненты достижения целевой ситуации, стоимости управляющего перехода и значений параметров следующей нечеткой ситуации. **Новизна**: реализуется подход, позволяющий использовать оценки управляющих решений не только для автономного выбора управляющих решений, но и для интеллектуальной поддержки действий оператора. Для каждого допустимого управляющего решения определяется оценка, характеризующая его ожидаемую полезность с учетом последующих переходов из ситуации в ситуацию. На основе полученных оценок формируется ранжированный набор альтернатив. **Результат**: описана процедура формирования ранжированного набора допустимых управляющих решений, согласованных с заданной стратегией управления. Рассмотрен пример поддержки принятия решений при управлении энергетической мини-сетью. Для альтернативных переходов сформированы числовые оценки, позволяющие обосновать выбор решений с учетом не только непосредственной стоимости их воздействия, но и ожидаемых последствий дальнейших решений. **Практическая значимость**: представленные результаты могут использоваться в системах поддержки принятия решений для операторов сложных технических и организационно-технических систем, предоставляя допустимые альтернативы решений, их ранжирование и количественные оценки ожидаемой полезности.*

***Ключевые слова**: поддержка принятия решений, ситуационное управление, оценка управляющих решений, адаптивные нечеткие ситуационные сети, глубокое обучение с подкреплением, Deep Recurrent Q-Network.*

Введение

Нечеткий ситуационный подход эффективно применим для поддержки принятия решений в условиях неполноты и неопределенности данных, с учетом выполнения требований и ограничений при переходе через промежуточные си-

Библиографическая ссылка на статью:
Шапкин А. П., Борисов В. В. Поддержка принятия решений на основе адаптивных нечетких ситуационных сетей и глубокого обучения с подкреплением // Системы управления, связи и безопасности. 2026. № 4. С. 88-102. DOI: 10.24412/2410-9916-2026-4-088-102

Reference for citation:
Shapkin A. P., Borisov V. V. Decision support based on adaptive fuzzy situational networks and deep reinforcement learning. *Systems of Control, Communication and Security*, 2026, no. 4, pp. 88-102 (in Russian). DOI: 10.24412/2410-9916-2026-4-088-102

туации для достижения целевой ситуации [1]. Одной из основных особенностей этого подхода является формирование модели рассматриваемой проблемы, совмещенной с моделью управления ею в виде совокупности управляющих решений для достижения целевой ситуации [2].

В системе поддержки принятия решений ситуационная модель должна использоваться не только для отображения текущего состояния объекта и множества допустимых переходов, но и для формирования рекомендаций лицу, принимающему решение. Такие рекомендации должны содержать перечень допустимых управляющих воздействий и их сравнительные оценки.

В работе [1] предложены нечеткие ситуационные сети, состоящие из нечетких ситуаций и дуг, соответствующих переходам между ситуациями при воздействии управляющих решений на соответствующие ситуационные признаки.

В статье [3] предложены нечеткие ситуационно-событийные модели, учитывающие неопределенность воздействия управляющих решений на ситуации за счет использования лингвистических лотерей, а также продолжительность выполнения решений.

Развитием этих моделей являются адаптивные нечеткие ситуационные модели, способы адаптации которых позволяют изменять их структуру и параметры в зависимости от изменения системных и внешних факторов, стратегий принятия решений и ограничений [4]. В работах [5, 6] предложены нечеткие ситуационно-прецедентные модели, учитывающие неоднозначность переходов в сетевой структуре между сложно связанными нечеткими ситуациями с учетом динамично изменяющихся системных и внешних факторов в условиях неопределенности. Однако определение наилучшей последовательности управляющих решений для перехода из текущей в целевую ситуацию и выбор следующей нечеткой ситуации в соответствии с заданной стратегией представляет собой нетривиальную задачу в условиях необходимости ее структурно-параметрической настройки. Подход к обучению с подкреплением создает хорошие предпосылки для решения указанной выше проблемы оценки управляющих решений при поиске наилучших их последовательностей с максимизацией функции полезности с учетом выбранной стратегии и постоянно накапливаемого опыта.

С позиции поддержки принятия решений требуется оценивать не только каждое доступное в текущей ситуации управляющее решение, но и их последовательность по ожидаемой полезности, допустимости и влиянию на достижение целевой ситуации.

1. Адаптивные нечеткие ситуационные сети

Адаптивная нечеткая ситуационная сеть (АНСС) может рассматриваться как формальная модель управляемой проблемной области, задаваемая взвешенным ориентированным графом. В такой структуре вершины интерпретируются как нечеткие ситуации, а дуги фиксируют допустимые переходы между ними, возникающие в результате применения управляющих решений к соответствующим ситуационным признакам. Описание каждой ситуации строится на осно-

ве нечетких множеств второго порядка, позволяющих представить состояние сложной системы через совокупность параметров и их лингвистических оценок. Существенным отличием АНСС от других ситуационных моделей является возможность структурно-параметрической адаптации. При изменении состава факторов, ограничений или условий функционирования системы модель может добавлять новые ситуации и управляющие переходы. За счет этого АНСС обеспечивает более полную картину мира при моделировании сложных систем в условиях неопределенности и изменчивости внешней среды [4].

Практической задачей при использовании АНСС является выбор наилучшей последовательности управляющих решений, которая позволит перевести систему из текущей ситуации в целевую с учетом выбранной стратегии управления. Для обеспечения следования стратегии управления необходимо оценить управляющие решения с учетом заданных критериев. В качестве таких критериев могут выступать минимизация числа управляющих воздействий, уменьшение суммарной стоимости переходов, повышение надежности или безопасности принимаемых решений, максимизация среднего веса маршрута как отношения суммарного веса переходов к их количеству, а также комбинированные показатели, отражающие требования конкретной задачи [2].

Выбор целевой ситуации и определение ближайшего управляющего воздействия осложняются тем, что сложная система функционирует в не полностью определенной и динамически изменяющейся среде. Поэтому заранее заданные правила выбора переходов не всегда позволяют получить решение, идеальное с точки зрения долгосрочных последствий. Для повышения обоснованности оценки последующих состояний целесообразно использовать методы обучения с подкреплением. В рамках АНСС они позволяют формировать количественные оценки управляющих решений на основе накопленного опыта взаимодействия с моделью и максимизации заданной функции полезности. Такой подход дает возможность сопоставлять альтернативные действия не только по локальным характеристикам перехода, но и по их вкладу в достижение целевой ситуации, что повышает гибкость и точность ситуационного управления сложными системами.

2. Обучение с подкреплением

Обучение с подкреплением относится к классу методов машинного обучения, в которых обучаемая система формирует поведение в процессе последовательного взаимодействия с внешней средой. Агент выполняет действия, наблюдает изменения состояния среды и получает награду за действия, приближающие его к цели, либо штраф за нежелательные последствия. В отличие от обучения по заранее размеченным примерам, здесь качество поведения определяется результатами собственных проб агента и накопленным опытом. Целью обучения с подкреплением является формирование политики поведения, которая обеспечивает максимизацию ожидаемой итоговой награды при взаимодействии со средой.

Формальным основанием большинства алгоритмов обучения с подкреплением является модель марковского процесса принятия решений (МППР). В этой модели среда в каждый момент времени находится в некотором состоянии, а выбранное агентом действие переводит ее в новое состояние и сопровождается соответствующим вознаграждением. Такое описание удобно для задач управления, поскольку позволяет связать локальные действия агента с долгосрочным результатом, выраженным через ожидаемую накопленную награду [7].

Структурно МППР и АНСС имеют несколько сходств. Обе модели описываются в виде взвешенных ориентированных графов. Вершинами в данном графе являются состояния системы или нечеткие ситуации для МППР и АНСС соответственно. Ребра графа соответствуют доступным переходам, которым соответствуют действия агента или управляющие решения. Поэтому формируемая в обучении с подкреплением политика поведения может интерпретироваться как аналог правил поведения, позволяющие выбирать дальнейшее управляющее воздействие в АНСС в соответствии с принятой стратегией ситуационного управления [8].

При поиске наилучших управляющих решений в АНСС некоторые алгоритмы обучения с подкреплением являются неприменимыми. Q-learning требует явного хранения таблицы значений для пар «состояние – действие», что плохо масштабируется при большом или изменяющемся пространстве ситуаций. DQN частично снимает это ограничение за счет аппроксимации Q-функции нейронной сетью, однако этот алгоритм ориентирован преимущественно на текущее наблюдение и не учитывает зависимость между последовательными переходами. Для АНСС это особенно существенно, поскольку предпочтительность действия может зависеть от уже пройденного маршрута, накопленных штрафов и изменений структуры модели [9, 10].

Алгоритмы глубокого обучения с подкреплением, например Deep Recurrent Q-Network (DRQN), расширяют идею DQN за счет включения рекуррентной нейронной сети, позволяющего использовать историю наблюдений при оценке последующих действий. LSTM-слои рекуррентной нейронной сети позволяют сохранять информацию о ранее выбранных управляющих решениях, что позволяет формировать более точные оценки управляющих решений [11]. Эти возможности делают данный алгоритм более подходящим для оценки действий в адаптивных нечетких ситуационных сетях, где структура модели может изменяться, а качество принятых решений определяется совокупностью различных факторов [12, 13].

3. Поддержка принятия решений на основе глубокой рекуррентной нейронной сети

3.1. Особенности обучения глубокой рекуррентной нейронной сети для оценки управляющих решений в АНСС

Алгоритмы глубокого обучения с подкреплением изначально разрабатывались для работы в условиях частичной наблюдаемости среды, что в контексте АНСС позволяет обученному агенту находить оптимальные пути при динами-

ческом расширении структуры сети. Важно подчеркнуть, что адаптация структуры происходит в строго заданных рамках, определяемых свойствами моделируемой сложной системы. В процессе обучения агент выявляет системные ограничения, формируя внутреннее представление о допустимых пространствах состояний и переходов.

Основным преимуществом использования алгоритмов обучения с подкреплением в контексте АНСС является возможность интеграции выбранной стратегии управления в процесс обучения через функцию вознаграждения $R(s, a, s')$. В эту функцию можно закодировать любые критерии, которые описывает стратегия управления: минимизация суммарного веса переходов, избегание нежелательных ситуаций, предпочтение вершин с высокими значениями некоторого параметра. Таким образом, обученная функция оценки $Q(s, a; \theta)$ позволяет получить оценку действий, отражающую долгосрочные последствия решений в контексте выбранной стратегии управления.

Обучение агента происходит на нескольких АНСС с различным количеством нечетких ситуаций и переходов, но в рамках одной предметной области. Такой подход обеспечивает обучение закономерностям присущим конкретному классу моделируемых сложных систем. Для этого строится начальная АНСС, которая затем достраивается новыми нечеткими ситуациями. Применяется методика обучения по учебному плану: сначала агент обучается на простой АНСС (10-12 нечетких ситуаций), затем сложность постепенно увеличивается (до 20-25 вершин) [14].

Для стабилизации обучения используется целевая сеть – копия основной глубокой рекуррентной нейронной сети с идентичной архитектурой, но обновляемая реже. Периодическая синхронизация параметров целевой сети с основной уменьшает корреляцию между текущими Q-значениями и целевыми, предотвращая расходимость [15].

3.2. Архитектура глубокой рекуррентной нейронной сети для оценки управляющих решений

Архитектура глубокой рекуррентной нейронной сети, используемая для оценки управляющих решений в АНСС, построена с учетом двух особенностей рассматриваемой задачи: необходимостью учитывать историю уже выполненных переходов и учет изменяемой структуры АНСС. Число нечетких ситуаций и переходов может изменяться, поэтому архитектура агента не должна быть жестко привязана к фиксированному числу вершин. Архитектура сети представлена на рис. 1.

Входной поток сети включает два типа данных. Первый тип данных соответствует наблюдению агента. Наблюдение формируется на основе признаков текущей и целевой нечетких ситуаций. При обучении такие наблюдения объединяются в последовательность фиксированной длины, что позволяет использовать рекуррентный слой для учета предшествующих переходов. При использовании обученного агента на каждом шаге подается текущее наблюдение, а

скрытое состояние LSTM сохраняется между шагами одного эпизода и сбрасываются при переходе к новому эпизоду [10–12].

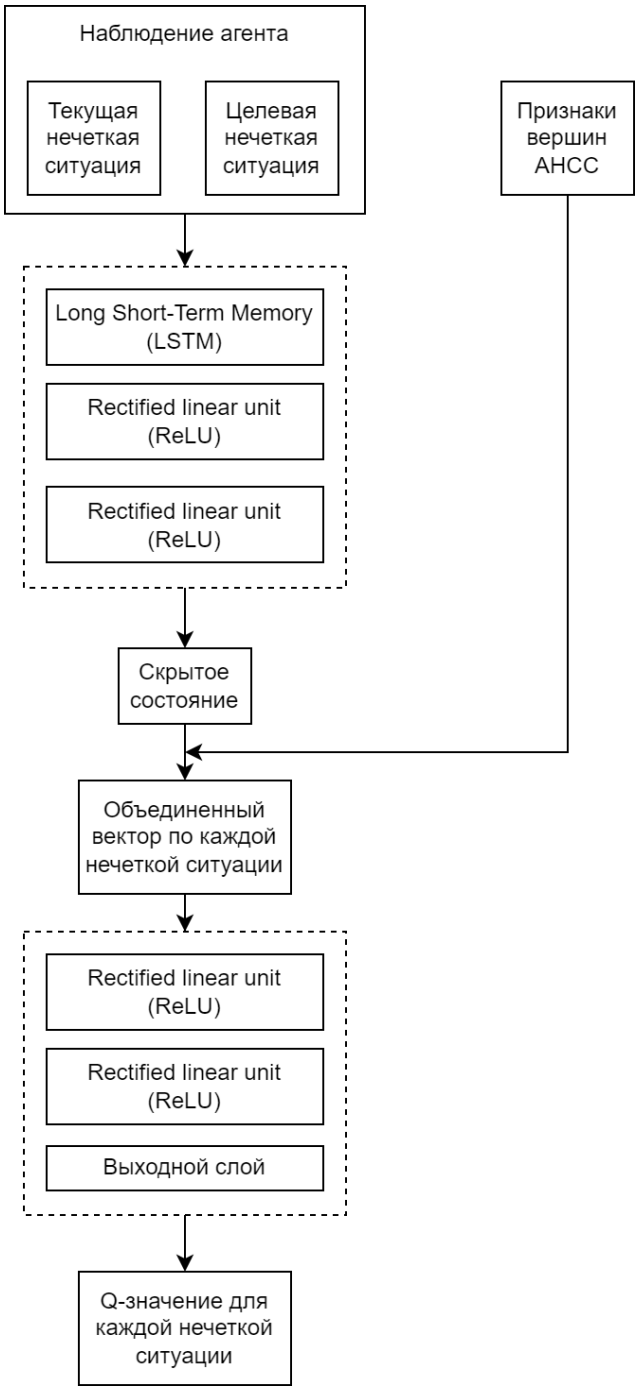


Рис. 1. Архитектура сети для оценки управляющих решений

Второй тип данных представляет собой матрицу признаков вершин АНСС. Каждая строка этой матрицы соответствует одной нечеткой ситуации и содержит значения функций принадлежности ее ситуационных признаков. Данная матрица позволяет оценивать потенциальный переход не только по текущему состоянию агента, но и по признакам каждой возможной следующей ситуации. В отличие от наблюдения агента, матрица признаков вершин не об-

рабатывается рекуррентным слоем, а используется на этапе вычисления Q-значений для всех потенциальных действий.

Последовательность наблюдений поступает на LSTM-слой. Его задача состоит в формировании скрытого состояния, отражающего контекст текущего маршрута: предшествующие переходы, ранее посещенные ситуации и накопленный опыт движения к целевой ситуации. После LSTM-слоя применяется нормализация слоя, а затем два полносвязных слоя с функцией активации ReLU, которая обнуляет отрицательные значения, поступившие на вход слоя. Эти слои формируют компактное признаковое представление состояния агента, используемое далее при оценке потенциальных переходов [11, 16, 17].

Полученное представление размножается по числу нечетких ситуаций и объединяется с матрицей признаков нечетких ситуаций АНСС. В результате для каждой нечеткой ситуации формируется объединенный вектор, дополнительно включающий информацию о посещенных нечетких ситуациях.

Далее каждый объединенный вектор подается на общую полносвязную сеть, применяемую ко всем вершинам с одними и теми же весами. Данная сеть выполняет роль оценочного модуля и состоит из двух скрытых полносвязных слоя с функцией активации ReLU и выходного слоя с одним нейроном. На выходе для каждой нечеткой ситуации формируется одно Q-значение. Данное значение интерпретируется как оценка ожидаемой полезности перехода в соответствующую ситуацию с учетом текущего наблюдения, истории маршрута и стратегии управления, заданной функцией вознаграждения [10, 17]. Количество выходных Q-значений определяется числом строк в матрице признаков вершин, а не фиксированным числом нейронов выходного слоя. Это позволяет применять один и тот же подход для АНСС различной размерности.

Для исключения невозможных управляющих решений используется маскирование действий. На этапе выбора действия из структуры АНСС определяется множество переходов, доступных из текущей нечеткой ситуации. После вычисления Q-значений для всех вершин значения, соответствующие недостижимым ситуациям, исключаются из выбора путем замены на заведомо малую величину. В результате агент выбирает действие только среди реально допустимых переходов. Это обеспечивает согласованность поведения агента с текущей структурой АНСС и предотвращает выбор переходов, отсутствующих в модели [15, 17].

3.3. Обучение глубокой рекуррентной нейронной сети для оценки управляющего решения

Процесс обучения агента начинается с инициализации параметров нейронной сети случайными значениями. Агент взаимодействует со средой АНСС, последовательно переходя из одной нечеткой ситуации в другую, выбираемую по ε-жадной стратегии. Данный переход сохраняется в буфер воспроизведения, применяемый для повышения устойчивости обучения. При обновлении параметров сети из этого буфера случайным образом выбираются эпизоды последовательностей переходов [18].

Обновление параметров глубокой рекуррентной нейронной сети происходит путем минимизации функции потерь между предсказанными Q-значениями и целевыми значениями, вычисленными с использованием отдельной целевой сети.

В контексте ситуационного подхода политика агента, полученная в результате обучения, представляет собой систему предпочтений, отражающую стратегию управления. Формально политика $\pi(a|s)$ определяет вероятность выбора действия a в ситуации s. В случае ε-жадной стратегии агент с вероятностью $1-\varepsilon$ выбирает действие с максимальным Q-значением, а с вероятностью ε исследует альтернативные варианты.

Стратегия управления, согласно которой действует агент формализуется в виде функции награды. Эта функция может описывать стратегию управления любой сложности: от простых, например, минимизация весов управляющего воздействия, так и более сложные стратегии, учитывающие значения нечетких параметров и условия эксплуатации моделируемой сложной системы. При этом веса управляющих решений напрямую не кодируются в массиве входных данных. Данные о стоимости управляющих воздействий интегрируются в функцию наград.

Для вычисления Q-значений в текущей нечеткой ситуации s_t входной массив признаков текущей ситуации формируется по значениям функций принадлежности всех признаков нечеткой ситуации. К этому вектору добавляется информация о целевой ситуации.

На каждом шаге эпизода в глубокую рекуррентную нейронную сеть подается текущее наблюдение и предыдущее скрытое состояние LSTM (изначально нулевое). Сеть вычисляет новые скрытые состояния и Q-значения для всех вершин. К выходному вектору применяется маска доступных из текущей ситуации переходов (получаемая из структуры АНСС). Таким образом, агент получает оценки только для реально возможных управляющих решений.

Скрытое состояние сохраняется между шагами одного эпизода, что позволяет учитывать контекст всей пройденной траектории. После завершения эпизода скрытое состояние сбрасывается, поскольку новый эпизод соответствует решению новой задачи (новый граф или новые начальная и целевая ситуации).

Полученные Q-значения интерпретируются как числовые оценки ожидаемой полезности каждого возможного перевода системы в следующую нечеткую ситуацию с учетом выбранной стратегии управления, заложенной в функцию вознаграждения. Чем выше Q-значение, тем более предпочтительным считается данное управляющее решение.

4. Применение глубокой рекуррентной нейронной сети для поддержки принятия решений

Рассмотрим пример оценки управляющих решений для адаптивной нечеткой ситуационной сети, моделирующей процесс управления энергетической мини-сетью [8]. Энергетическая мини-сеть рассматривается как автономная или локальная система электроснабжения, включающая возобновляемый источник

генерации, аккумуляторную батарею, резервный генератор и потребителей электроэнергии. В процессе функционирования такой системы необходимо поддерживать баланс между выработкой и потреблением энергии, обеспечивать допустимый уровень заряда аккумуляторов, снижать использование резервного генератора и не допускать перехода оборудования в нежелательные режимы.

Множество признаков $P = \{p_1, p_2, p_3, p_4\}$, где p_1 – уровень заряда аккумуляторов, p_2 – производство энергии солнечными панелями, p_3 – потребление энергии, p_4 – количество топлива в резервном генераторе. Нечеткие параметры сформированы на основе основных измеряемых или расчетных характеристик мини-сети: p_1 – отражает доступный запас энергии и возможность покрытия дефицита генерации; p_2 – характеризует текущий вклад возобновляемого источника; p_3 – задает нагрузку, которую необходимо обеспечить; p_4 – отражает ресурс для поддержания электроснабжения при недостаточной генерации от возобновляемых источников. Для каждого из указанных признаков вводятся лингвистические значения, например, «низкий», «средний» и «высокий», что позволяет представить состояние мини-сети в виде нечеткой ситуации. Структура АНСС на момент решения задачи представлена на рис. 2.

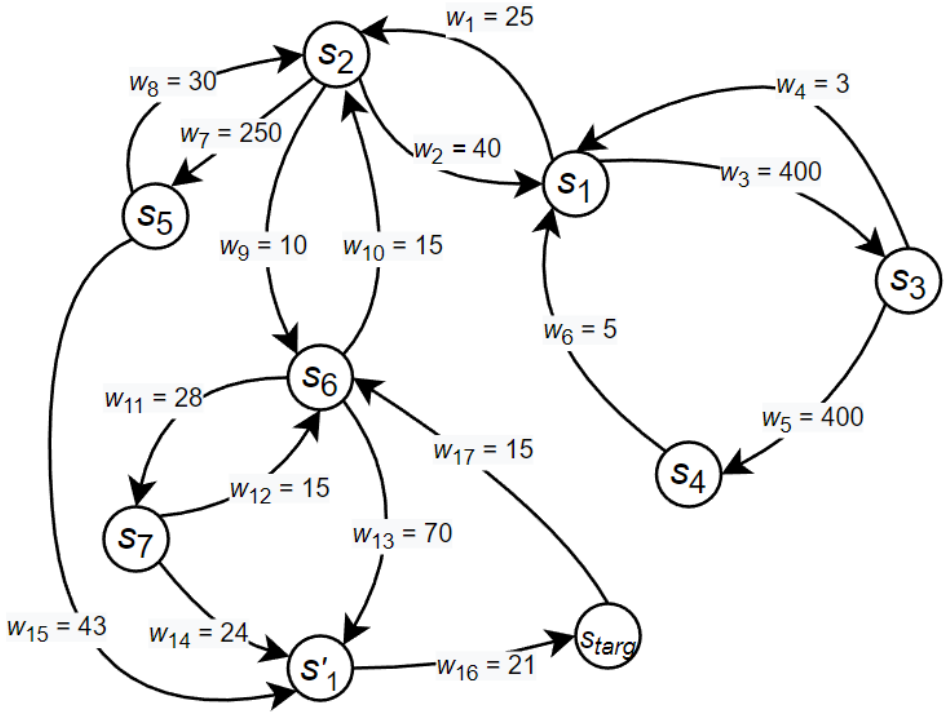


Рис. 2. АНСС энергетической мини-сети

Требуется перевести систему в целевую нечеткую ситуацию s_{targ}. В данный момент система находится в ситуации s_6, ранее она была в ситуации s_2. Стратегия управления предполагает минимизацию суммарной стоимости управляющих решений, но при этом некоторые переходы считаются нежелательными из-за повышенных рисков для оборудования. В частности, прямой переход в s'_1 (обозначим его как действие a_1) имеет высокую стоимость (большой вес ребра), но является безопасным. Альтернативный путь через ситуацию

s_7 (действие a_2) имеет меньшую стоимость, однако эксплуатация системы в состоянии s_7 сопряжена с повышенными рисками, поэтому такие переходы штрафуются в функции вознаграждения. Функция вознаграждения формализует данную стратегию следующим образом:

– за каждый шаг начисляется отрицательная величина, пропорциональная весу перехода;
– при достижении целевой ситуации добавляется положительная награда;
– за посещение нежелательных ситуаций (например, s_7) добавляется дополнительный отрицательный штраф.

В результате обучения глубокая рекуррентная нейронная сеть формирует Q-функцию, которая отражает предпочтения, заданные стратегией. Для текущей ситуации s_6 агент, получив на вход признаки ситуации и скрытое состояние, аккумулирующее историю предыдущих переходов (в частности, тот факт, что система пришла из s_2), вычисляет Q-значения для доступных действий. После применения маски доступных переходов получаются следующие оценки:

– $Q(s_6, a_1) = 15{,}8$ (переход непосредственно в s'_1);
– $Q(s_6, a_2) = 12{,}7$ (переход в s_7).

Более высокое значение для прямого перехода в s'_1 соответствует стратегии, избегающей рискованного состояния s_7, несмотря на его меньшую стоимость. Таким образом, обученный агент предоставляет количественную оценку каждого возможного управляющего решения, согласованную с заданными критериями оптимальности.

Заключение

В статье реализован подход к поддержке принятия решений на основе адаптивных нечетких ситуационных сетях и глубокого обучения с подкреплением. При этом выбор последовательности управляющих решений выполняется с учетом текущей и целевой ситуаций, истории ранее выполненных переходов, ограничений структуры модели и заданной стратегии управления.

Для оценки управляющих решений за основу взят алгоритм глубокого обучения с подкреплением. LSTM-слои рекуррентной нейронной сети позволяют формировать скрытое состояние с историей пройденных нечетких ситуаций. Скрытое состояние объединяется с матрицей признаков каждой нечеткой ситуации и с помощью полносвязной нейронной сети формируются оценки доступных управляющих решений. Динамическое маскирование исключает из сформированного набора управляющие решения, которым не соответствует реальный переход из текущей ситуации.

На примере управления энергетической мини-сетью проиллюстрированы процедуры оценки и формирования последовательностей наиболее эффективных управляющих решений, согласованные с заданной стратегией управления.

Исследование выполнено в рамках государственного задания, проект № FSWF 2026-0010.

Литература

1. Мелихов А. Н., Берштейн Л. С., Коровин С. Я. Ситуационные советующие системы с нечеткой логикой. М.: Наука, 1990. 272 с.

2. Борисов В. В., Авраменко Д. Ю. Нечеткое ситуационное управление сложными системами на основе их композиционного гибридного моделирования // Системы управления, связи и безопасности. 2021. № 3. С. 207-237. DOI: 10.24412/2410-9916-2021-3-207-237.

3. Борисов В. В., Зернов М. М. Реализация ситуационного подхода на основе нечеткой иерархической ситуационно-событийной сети // Искусственный интеллект и принятие решений. 2009. № 1. С. 17-30.

4. Denisenkov M. A., Borisov V. V. Modeling the behavior of intelligent agents based on adaptive fuzzy situational networks // Proceedings of the 3rd Russian-Pacific Conference on Computer Technology and Application (RPC). Vladivostok, Russia, 18-25 Aug. 2018. P. 1–5. DOI: 10.1109/RPC.2018.8482217.

5. Соколов А. М., Борисов В. В. Рекуррентные нечеткие ситуационно-прецедентные модели для оперативного управления сложными техническими объектами по прецедентам // Программные продукты и системы. 2024. Т. 37. № 4. С. 554–565. DOI: 10.15827/0236-235X.148.554-565.

6. Дли М. И., Черновалова М. В., Соколов А. М. Построение нечетких ситуационно-прецедентных моделей региональных экономических систем с использованием онтологий // Прикладная информатика. 2025. Т. 20. № 2. С. 112-125. DOI: 10.37791/2687-0649-2025-20-2-112-125.

7. AlMahamid F., Grolinger K. Reinforcement Learning Algorithms: An Overview and Classification // arXiv.org. 2022. URL: https://arxiv.org/abs/2209.14940 (дата обращения: 16.03.2024).

8. Борисов В. В., Шапкин А. П. Нечеткие ситуационные сети с адаптацией на основе обучения с подкреплением // Вестник Московского энергетического института. 2025. № 3. С. 135-143. DOI: 10.24160/1993-6982-2025-3-135-143.

9. Шапкин А. П., Борисов В. В. Способы внедрения обучения с подкреплением в адаптивные нечеткие ситуационные сети // XXII Международная научно-техническая конференция студентов и аспирантов (Смоленск, 23-24 апреля 2025 г.). Труды конференции. В 2 т. Т. 1. Смоленск: Универсум, 2025. С. 470-474.

10. Mnih V., Kavukcuoglu K., Silver D., Rusu A. A., Veness J., Bellemare M. G., Graves A., Riedmiller M., Fidjeland A. K., Ostrovski G., Petersen S., Beattie C., Sadik A., Antonoglou I., King H., Kumaran D., Wierstra D., Legg S., Hassabis D. Human-level control through deep reinforcement learning // Nature. 2015. Vol. 518. P. 529–533. DOI: 10.1038/nature14236.

11. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9, no. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

12. Hausknecht M., Stone P. Deep Recurrent Q-Learning for Partially Observable MDPs // arXiv.org. 2015. URL: https://arxiv.org/abs/1507.06527 (дата обращения: 16.03.2024).

13. Moreno-Vera F. Performing Deep Recurrent Double Q-Learning for Atari Games // arXiv.org. 2019. URL: https://arxiv.org/abs/1908.06040 (дата обращения: 16.03.2024).

14. Narvekar S., Peng B., Leonetti M., Sinapov J., Taylor M. E., Stone P. Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey // Journal of Machine Learning Research. 2020. Vol. 21, no. 181. P. 1–50.

15. Huang S., Ontañón S. A Closer Look at Invalid Action Masking in Policy Gradient Algorithms // arXiv.org. 2020. URL: https://arxiv.org/abs/2006.14171 (дата обращения: 16.03.2024).

16. Ba J. L., Kiros J. R., Hinton G. E. Layer Normalization // arXiv.org. 2016. URL: https://arxiv.org/abs/1607.06450 (дата обращения: 22.06.2026).

17. Dulac-Arnold G., Evans R., van Hasselt H., Sunehag P., Lillicrap T., Hunt J., Mann T., Weber T., Degris T., Coppin B. Deep Reinforcement Learning in Large Discrete Action Spaces // arXiv.org. 2015. URL: https://arxiv.org/abs/1512.07679 (дата обращения: 22.06.2026).

18. Zhu P., Li X., Poupart P., Miao G. On Improving Deep Reinforcement Learning for POMDPs // arXiv.org. 2018. URL: https://arxiv.org/abs/1804.06309 (дата обращения: 16.03.2024).

References

1. Melikhov A. N., Bershteyn L. S., Korovin S. Ya. *Situatsionnye sovetuyushchie sistemy s nechetkoy logikoy* [Situational Advising Systems with Fuzzy Logic]. Moscow, Nauka Publ., 1990. 272 p. (in Russian).

2. Borisov V. V., Avramenko D. Yu. Fuzzy Situational Control of Complex Systems Based on Composite Hybrid Modeling. *Systems of Control, Communication and Security*, 2021, no. 3, pp. 207–237, DOI: 10.24412/2410-9916-2021-3-207-237 (in Russian).

3. Borisov V. V., Zernov M. M. Realizatsiya situatsionnogo podkhoda na osnove nechetkoy ierarkhicheskoy situatsionno-sobytiynoy seti [Implementation of the Situational Approach Based on a Fuzzy Hierarchical Situational-Event Network]. *Artificial Intelligence and Decision Making*, 2009, no. 1, pp. 17–30 (in Russian).

4. Denisenkov M. A., Borisov V. V. Modeling the Behavior of Intelligent Agents Based on Adaptive Fuzzy Situational Networks. *Proceedings of the 3rd Russian-Pacific Conference on Computer Technology and Application (RPC)*, Vladivostok, Russia, 18–25 August 2018, pp. 1–5, DOI: 10.1109/RPC.2018.8482217.

5. Sokolov A. M., Borisov V. V. Recurrent Fuzzy Situational-Precedent Models for Operational Control of Complex Technical Objects. Programmnye produkty i sistemy, 2024, vol. 37, no. 4, pp. 554–565, DOI: 10.15827/0236-235X.148.554-565 (in Russian).

6. Dli M. I., Chernovalova M. V., Sokolov A. M. Construction of Fuzzy Situational-Precedent Models of Regional Economic Systems Using Ontologies. *Journal of Applied Informatics*, 2025, vol. 20, no. 2, pp. 112–125, DOI: 10.37791/2687-0649-2025-20-2-112-125 (in Russian).

7. AlMahamid F., Grolinger K. Reinforcement Learning Algorithms: An Overview and Classification. arXiv, 2022. Available at: https://arxiv.org/abs/2209.14940 (accessed 16 March 2024).

8. Borisov V. V., Shapkin A. P. Fuzzy Situational Networks with Reinforcement Learning-Based Adaptation. *MPEI Vestnik*, 2025, no. 3, pp. 135–143, DOI: 10.24160/1993-6982-2025-3-135-143 (in Russian).

9. Shapkin A. P., Borisov V. V. Sposoby vnedreniya obucheniya s podkrepleniem v adaptivnye nechetkie situatsionnye seti [Methods for Integrating Reinforcement Learning into Adaptive Fuzzy Situational Networks]. *XXII Mezhdunarodnaya nauchno-tekhnicheskaya konferentsiya studentov i aspirantov. Trudy konferentsii* [Proceedings of the 22nd International Scientific and Technical Conference of Students and Postgraduates], Smolensk, 23–24 April 2025, Smolensk, Universum Publ., 2025, vol. 1, pp. 470–474 (in Russian).

10. Mnih V., Kavukcuoglu K., Silver D., Rusu A. A., Veness J., Bellemare M. G., Graves A., Riedmiller M., Fidjeland A. K., Ostrovski G., Petersen S., Beattie C., Sadik A., Antonoglou I., King H., Kumaran D., Wierstra D., Legg S., Hassabis D. Human-Level Control through Deep Reinforcement Learning. *Nature*, 2015, vol. 518, pp. 529–533, DOI: 10.1038/nature14236.

11. Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation*, 1997, vol. 9, no. 8, pp. 1735–1780, DOI: 10.1162/neco.1997.9.8.1735.

12. Hausknecht M., Stone P. Deep Recurrent Q-Learning for Partially Observable MDPs. arXiv, 2015. Available at: https://arxiv.org/abs/1507.06527 (accessed 16 March 2024).

13. Moreno-Vera F. Performing Deep Recurrent Double Q-Learning for Atari Games. arXiv, 2019. Available at: https://arxiv.org/abs/1908.06040 (accessed 16 March 2024).

14. Narvekar S., Peng B., Leonetti M., Sinapov J., Taylor M. E., Stone P. Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey. *Journal of Machine Learning Research*, 2020, vol. 21, no. 181, pp. 1–50.

15. Huang S., Ontañón S. A Closer Look at Invalid Action Masking in Policy Gradient Algorithms. arXiv, 2020. Available at: https://arxiv.org/abs/2006.14171 (accessed 16 March 2024).

16. Ba J. L., Kiros J. R., Hinton G. E. Layer Normalization. arXiv, 2016. Available at: https://arxiv.org/abs/1607.06450 (accessed 22 June 2026).

17. Dulac-Arnold G., Evans R., van Hasselt H., Sunehag P., Lillicrap T., Hunt J., Mann T., Weber T., Degris T., Coppin B. Deep Reinforcement Learning in Large Discrete Action Spaces. arXiv, 2015. Available at: https://arxiv.org/abs/1512.07679 (accessed 22 June 2026).

18. Zhu P., Li X., Poupart P., Miao G. On Improving Deep Reinforcement Learning for POMDPs. arXiv, 2018. Available at: https://arxiv.org/abs/1804.06309 (accessed 16 March 2024).

Статья поступила 26 августа 2026 г.
Одобрена после рецензирования 10 апреля 2026 г.
Принята к публикации 21 апреля 2026 г.

Информация об авторах

Шапкин Александр Павлович – аспирант кафедры вычислительной техники. Филиал федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет «МЭИ» в г. Смоленске. Область научных интересов: интеллектуальный анализ; моделирование сложных систем и процессов. E-mail: shapkinap1999@gmail.com

Борисов Вадим Владимирович – доктор технических наук, профессор. Заведующий кафедры вычислительной техники. Филиал федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет «МЭИ» в г. Смоленске. Старший научный сотрудник научно-исследовательского центра. Военная академия войсковой противовоздушной обороны Вооруженных сил Российской Федерации им. А.М. Василевского. Области научных интересов: нечеткий и нейро-нечеткий анализ, моделирование сложных систем и процессов; интеллектуальная поддержка принятия решений; ассоциативные системы хранения и обработки информации. E-mail: vbor67@mail.ru

Адрес: 214013, Россия, г. Смоленск, Энергетический проезд, д. 1.

Decision support based on adaptive fuzzy situational networks and deep reinforcement learning

A. P. Shapkin, V. V. Borisov

***Problem statement**: To support control decision-making in complex technical systems operating under uncertainty, it is necessary to determine the possible system states, available control actions, and a control strategy. Known fuzzy situational models allow one to represent a complex system as a network of situations and control actions; however, they do not provide automatic evaluation of permissible control decisions taking into account the selected control strategy. **Objective**: to develop a method for supporting control decision-making in adaptive fuzzy situational networks based on deep reinforcement learning, ensuring the evaluation of permissible control actions taking into account their long-term consequences. **Methods used**: the process of generating recommendations is presented as a reinforcement learning problem. The deep reinforcement learning method is used to evaluate control decisions. A target network and an experience replay buffer are used to stabilize learning. Inadmissible actions are excluded using a binary transition mask generated based on the current structure of the adaptive fuzzy situational network. The control strategy is defined by a parameterized reward function including components of achieving the target situation, the cost of the control transition, and the values of the parameters of the next fuzzy situation. **Novelty**: a method is proposed that enables the use of control decision estimates not only for autonomous action selection but also for intelligent operator support. For each feasible control decision, an estimate is determined that characterizes its expected utility, taking into account subsequent transitions. Based on the obtained estimates, a ranked set of alternatives is formed. Joint processing of the observation history and features of all fuzzy situations allows for the application of a single agent architecture to networks of varying dimensions, while dynamic masking ensures the compliance of recommendations with the current model structure. **Result**: a method for generating a ranked set of feasible control decisions consistent with a given control strategy has been developed. An example of decision support for managing an energy mini-grid is considered. Numerical estimates are generated for alternative transitions, allowing for the justification of the choice of a safe solu-*

tion, taking into account not only the immediate cost of the impact but also the expected consequences of further movement through the network. ***Practical significance****: The proposed method can be used in intelligent decision support systems for operators of complex technical and organizational systems. The system provides the operator with feasible alternatives, their ranking, and quantitative estimates of expected utility, while the final selection and implementation of the control decision remains with the decision maker.*

Key words: *decision support, situational management, control decision evaluation, adaptive fuzzy situational networks, deep reinforcement learning, Deep Recurrent Q-Network.*

Information about Authors

Alexander Pavlovich Shapkin – Doctoral Student. The postgraduate student of the Department of Computer Engineering. Branch of the National Research University Moscow Power Engineering Institute in Smolensk. Field of research: intellectual analysis; modeling of complex systems and processes. E-mail: shapkinap1999@gmail.com

Vadim Vladimirovich Borisov – Dr. habil. of Engineering Sciences, Professor. Professor of the Department of Computer Engineering. Branch of the National Research University Moscow Power Engineering Institute in Smolensk. Senior researcher. Military Academy of Army Air Defence A.M. Vasilevsky. Field of research: fuzzy and fuzzy neural models and networks, intellectual decision-making support, associative memory, associative systems of storage and processing of the information and knowledge. E-mail: vbor67@mail.ru

Address: 214013, Russia, Smolensk, Energeticheskiy proezd, 1.